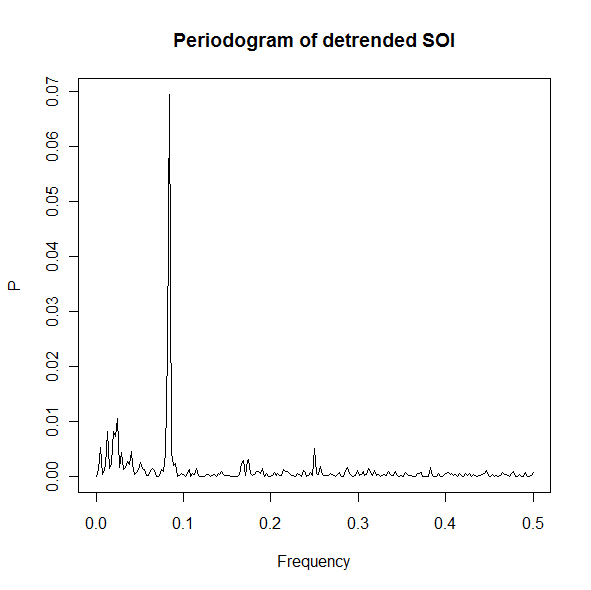
**Stat 510 Week 7 Homework**

**Submit answers in a Word document or a pdf to the designated ANGEL Drop Box in the Drop Boxes folder.**

1. Following is the periodogram of the detrended SOI series that we’ve used a few times in the past. SOI = southern oscillation index, a measure of temperature in the southern hemisphere. The series is n = 452 monthly values. The detrended values are the residuals from a linear regression on time (1 to 452).



**a**. Roughly, at what frequency is the dominant peak? The period of a cycle is the reciprocal of the frequency. About what is the period for the frequency at which the dominant peak occurs? (Keep in mind that this is monthly data and this is a weather measure.)

**b**. Roughly what is the region of frequency range for the second highest peak(s)? Roughly what would be the period(s) associated with these frequencies? (Hint: The El Nino weather phenomenon is believed to affect things about every 3 to 7 years.)

**2.** Do problem 4.1 on page 255 of the textbook. For part a, use a modified version of the R code on page 178. To simply increase the sample size from 100 to 128, you need only change 1:100 to 1:128. For part b, use a modification of the code on page 179 or the code below. To print the periodogram between 0 and .5 only, as opposed to between 0 and 1 in Figure 4.2, you may use the following code:

P = abs(2\*fft(x)/128)^2

f = 0:64/128

plot(f, P[1:65], type="o", xlab="frequency", ylab="periodogram")

For part c, it’s only necessary to plot the series that is the sum and plot the periodogram of the sum. The command to get the sum will be something like sum = x1+x2+x3+rnorm(100,0,5). The parameters for the rnorm command are sample size, mean, standard deviation. (The notation the authors used in part c gives the mean and the variance.)

For answers to the book parts–

Part a Give plots of the x1, x2, x3 and the sum. To determine how things differ from Example 4.1, focus on the formula for the series rather than the plots.

Part b. Give the periodogram of the sum and describe the main frequency peaks. What has changed from Figure 4.2?

Part c. Give the plot of the sum and its periodogram. And, describe the locations of the main peaks in the periodogram.

**3.** Use the dataset “speech.dat” from the Week 6 folder. The data are described in Example 1.3 on page 5 of the textbook. The series is a sample of recorded speech of the phrase aaa…hhh, sampled for 0.1020 seconds at the rate of 10,000 points per second. Thus there are n = 1020 data values. The authors don’t say what exactly is the response variable, but a likely possibility for the response is a measure of the audio frequency of the sound. (Maybe they held out from saying that to avoid confusion with the frequency of a cosine wave.)

The data can be read as x = scan("speech.dat") once the data file is in place on your computer.

There’s no obvious trend seen in Figure 1.3 on page 6, but it never hurts to detrend before looking at a periodogram. At worst, detrending does nothing. Detrend in R by doing a linear regression on the time index. The detrended values are the residuals from this regression. The sequence of commands might be something like

t = 1:1020

regr = lm(x~t)

y = residuals(regr)

In the above code, y would be the detrended series, the residuals from a linear regression.

Plot the periodogram of the detrended series. Give the plot and describe the pattern of peaks as they relate to frequency (the horizontal axis).

**4.** This problem is a review of the material in Week 3 (non-seasonal ARIMA models). Use the lakesim100.dat dataset in the Week 6 folder. It gives *n* =100 consecutive annual measures of the level of a lake.

a. Plot the time series. Give the plot and an interpretation of the plot as the answer to this part. Please review Lesson 1.1 for what you should interpret in the plot.

b. Determine the ACF and PACF for the series. (Hint: The most efficient way to do this is to load the astsa library and then use the acf2 command.)

Give the ACF and PACF plots AND write a brief interpretation. Do you think the data are stationary? Do you think you should suggest models from the ACF and PACF?

c. Give the ACF and PACF of first differences. Write a brief interpretation that indicates what model(s) you think are suggested.

d. Determine an ARIMA model that you think is suitable for these data. Describe the model and give the estimated coefficients. (You can just give the output that you get from R.)

e. For your model, give diagnostic plots/tests for checking the suitability of the model (for example, the diagnostic output from the sarima command). Briefly discuss what the plots indicate about your model.